PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/54, 1/21, 9/12, C12Q 1/68, C07K 16/40, A61K 48/00, 38/43, G01N 33/53

A1

(11) International Publication Number: WO 97/21819

(43) International Publication Date: 19 June 1997 (19.06.97)

(21) International Application Number: PCT/GB96/03047

(22) International Filing Date: 11 December 1996 (11.12.96)

(30) Priority Data:

9525246.6	11 December 1995 (11.12.95)	GB
9604580.2	4 March 1996 (04.03.96)	GB
9604581.0	4 March 1996 (04.03.96)	GB
9612887.1	20 June 1996 (20.06.96)	GB
60/027,972	8 October 1996 (08.10.96)	US

(71) Applicant *(for all designated States except US)*: SMITHKLINE BEECHAM PLC [GB/GB]; New Horizons Court, Brentford, Middlesex TW8 9EP (GB).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: WALLIS, Nicola [GB/US]; SmithKline Beecham Pharmaceuticals, 1250 South Collegeville Road, P.O. Box 5089, Collegeville, PA 19426-0989 (US). HODGSON, John, Edward [GB/US]; SmithKline Beecham Pharmaceuticals, 1250 South Collegeville Road, P.O. Box 5089, Collegeville, PA 19426-0989 (US). BURNHAM, Martin, Karl, Russell [GB/US]; SmithKline Beecham Pharmaceuticals, 1250 South Collegeville Road, P.O. Box 5089, Collegeville, PA 19426-0989 (US).

(74) Agent: GIDDINGS, Peter, John; SmithKline Beecham, Corporate Intellectual Property, Two New Horizons Court, Brentford, Middlesex TW8 9EP (GB).

(81) Designated States: JP, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: NOVEL NAGPU

(57) Abstract

Novel NAGPU polypeptides and DNA (RNA) encoding such novel protein and a procedure for producing such polypeptides by recombinant techniques is disclosed. Also disclosed are methods for utilizing this novel protein for the treatment of infection, particularly bacterial infections. Antagonists against such novel protein and their use as a therapeutic to treat infections, particularly bacterial infections are also disclosed. Also disclosed are diagnostic assays for detecting diseases related to the presence of novel protein encoding nucleic acid sequences and the polypeptides in a host. Also disclosed are diagnostic assays for detecting polynucleotides encoding protein of the novel protein's family and for detecting the polypeptide in a host.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	**GB**	United Kingdom	**MW**	Malawi
AT	Austria	**GE**	Georgia	**MX**	Mexico
AU	Australia	**GN**	Guinea	**NE**	Niger
BB	Barbados	**GR**	Greece	**NL**	Netherlands
BE	Belgium	**HU**	Hungary	**NO**	Norway
BF	Burkina Faso	**IE**	Ireland	**NZ**	New Zealand
BG	Bulgaria	**IT**	Italy	**PL**	Poland
BJ	Benin	**JP**	Japan	**PT**	Portugal
BR	Brazil	**KE**	Kenya	**RO**	Romania
BY	Belarus	**KG**	Kyrgystan	**RU**	Russian Federation
CA	Canada	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CF	Central African Republic			**SE**	Sweden
CG	Congo	**KR**	Republic of Korea	**SG**	Singapore
CH	Switzerland	**KZ**	Kazakhstan	**SI**	Slovenia
CI	Côte d'Ivoire	**LI**	Liechtenstein	**SK**	Slovakia
CM	Cameroon	**LK**	Sri Lanka	**SN**	Senegal
CN	China	**LR**	Liberia	**SZ**	Swaziland
CS	Czechoslovakia	**LT**	Lithuania	**TD**	Chad
CZ	Czech Republic	**LU**	Luxembourg	**TG**	Togo
DE	Germany	**LV**	Latvia	**TJ**	Tajikistan
DK	Denmark	**MC**	Monaco	**TT**	Trinidad and Tobago
EE	Estonia	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**UG**	Uganda
FI	Finland	**ML**	Mali	**US**	United States of America
FR	France	**MN**	Mongolia	**UZ**	Uzbekistan
GA	Gabon	**MR**	Mauritania	**VN**	Viet Nam

NOVEL NAGPU

FIELD OF THE INVENTION

This invention relates to newly identified polynucleotides, polypeptides encoded by such polynucleotides, the use of such polynucleotides and polypeptides, as well as the production of such polynucleotides and polypeptides and recombinant host cells transformed with the polynucleotides. The invention also relates to inhibiting the action of such polypeptides and to the use of inhibitors in therapy.

BACKGROUND OF THE INVENTION

N-Acetylglucosamine 1-Phosphate Uridyltransferase (NAGPU) catalyzes the formation of UDP-N-Acetyl Glucosamine, an essential precursor for cell wall peptidoglycan in all bacteria and of lipopolysaccharide and enterobacterial common antigen in gram-negatives. The enzyme has been purified from *Escherichia coli* and is bifunctional, also catalyzing the preceding step of N-acetylation of glucosamine-1-phosphate (Mengin-Lecreulx, D. and van Heijenoort, J, J.Bacteriol. 176: 5788-5795 [1994]). It is possible to block the acetyltransferase activity but not the uridyl transferase activity with thiol inhibitors, suggesting that the enzyme may have two domains. The gene glmU encoding the enzyme has been cloned from *E. coli* (Mengin-Lecreulx, D. and van Heijenoort, J, J.Bacteriol. 175: 6150-6157 [1993]) and its counterpart in *Bacillus subtilis* (gcaD) has also been identified (Hove-Jensen B, J.Bacteriol. 174: 6852-6 [1992]).

The essential nature of the gene product of gcaD is demonstrated by temperature sensitive mutants of *Bacillus subtilis* which are unable to make active enzyme and stop growing at the restrictive temperature (Hove-Jensen [1992]). The discovery of the gene from the human pathogen *Staphylococcus aureus* corresponding to gcaD permits one to produce NAGPU enzyme which can be used to screen for novel antibiotics as described below.

Recently, several novel approaches have been described which purport to follow global gene expression during infection (Chuang, S. et al. [1993] Global Regulation of Gene Expression in *Escherichia coli* J. Bacteriol. 175, 2026-2036, Mahan, M.J. et al. [1993] Selection of Bacterial Virulence Genes That Are Specifically Induced in Host Tissues SCIENCE 259, 686-688, Hensel, M. et al. [1995] Simultaneous Identification of Bacterial Virulence Genes by Negative Selection SCIENCE 269, 400-403). These new techniques have so far been demonstrated with gram negative pathogen infections and not with infections with gram positives presumably due to the much slower development of global transposon mutagenesis and suitable vectors needed for these strategies in these organisms. and in the case of that process described by Chuang, S. et al.[1993] the difficulty of isolating suitable quantities of bacterial RNA free of mammalian RNA derived from the infected tissue to furnish bacterial RNA labeled to sufficiently high specific

activity. The present invention employs a novel technology to determine gene expression in the pathogen at different stages of infection of the mammalian host. A novel aspect of this invention is the use of a suitably labeled oligonucleotide probe which anneals specifically to the bacterial ribosomal RNA in Northern blots of bacterial RNA preparations from infected tissue. Using the more abundant ribosomal RNA as a hybridization target greatly facilitates the optimization of a protocol to purify bacterial RNA of a suitable size and quantity for RT-PCR from infected tissue.

A suitable oligonucleotide useful for applying this method to genes expressed in *Staphylococcus aureus* is, for example, 5'-gctcctaaaaggttactccaccggc-3' [SEQ ID NO:6].

Use of the technology of the present invention enables identification of bacterial genes transcribed during infection, inhibitors of which would have utility in anti-bacterial therapy. Specific inhibitors of such gene transcription or of the subsequent translation of the resultant mRNA or of the function of the corresponding expressed proteins would have utility in anti-bacterial therapy.

SUMMARY OF THE INVENTION

The present invention relates to a novel protein from *S. aureus* WCUH 29, characterized in that it comprises an amino acid sequence selected from the group consisting of the amino acid sequence set forth in SEQ ID NO:2 and SEQ ID NO:5, or a fragment, analogue or derivative thereof either. This protein and any variants thereof, as well as polynucleotides encoding the same are herein referred to as "NAGPU."

The invention also relates to a polypeptide fragment of the protein, having the amino acid sequence selected from the group consisting of the amino acid sequence set forth in SEQ ID NO:2 and SEQ ID NO:5, or a derivative thereof.

In accordance with another aspect of the present invention, there are provided polynucleotides (DNA or RNA) which encode such polypeptides.

In particular the invention provides a polynucleotide having the DNA sequence given in SEQ ID NO:1, and a complementary sequence thereto.

The invention also provides an isolated polynucleotide comprising a member selected from the group consisting of: a polynucleotide having at least a 70% identity to a polynucleotide encoding a polypeptide comprising amino acids 1 to 450 of SEQ ID NO:2; a polynucleotide having at least a 70% identity to a polynucleotide encoding a polypeptide comprising amino acids 1 to 452 of SEQ ID NO:5; a polynucleotide which is complementary to the polynucleotide of any such isolated polynucleotide; and a polynucleotide comprising at least 15 sequential bases of the polynucleotide of any of such isolated polynucleotides.

Further provided is a polynucleotide comprising a polynucleotide sequence selected from the group consisting of the complementary sequence to nucleotide 54 to 1406 set forth in SEQ ID NO:1, and a complementary sequence to nucleotide 54 to 1412 set forth in SEQ ID NO:1.

In particular the invention provides a polynucleotide which encodes a polypeptide comprising amino acid selected from the group consisting of amino acid 1 to 450 of SEQ ID NO:2 and amino acid 1 to 452 of SEQ ID NO:5.

The invention provides an isolated polynucleotide comprising a member selected from the group consisting of: a polynucleotide having at least a 70% identity to a polynucleotide encoding the same mature polypeptide expressed by the DNA contained in NCIMB Deposit No. 40794 and having the polynucleotide sequence of SEQ ID NO:1; a polynucleotide complementary to the polynucleotide of any such isolated polynucleotide; and a polynucleotide comprising at least 15 bases of the polynucleotide of such isolated polynucleotides.

Also provided by the invention is a polypeptide comprising an amino acid sequence which is at least 70% identical to amino acid selected from the group consisting of 1 to 450 of SEQ ID NO:2 and amino acid 1 to 452 of SEQ ID NO:5.

Further provided by the invention is a polypeptide comprising an amino acid sequence selected from the group consisting of the amino acids as set forth in SEQ ID NO:2 and the amino acids as set forth in SEQ ID NO:5.

The present invention also provides a novel protein from *Staphylococcus. aureus* WCUH29 obtainable by expression of a gene characterized in that it comprises the DNA sequence given SEQ ID NO:1, or a fragment, analogue or derivative thereof.

The invention also relates to novel oligonucleotides, including those set forth in SEQ ID NO: 3, 4 and 6, SEQ ID NO:3 and 4 of which are derived from the polynucleotide sequence of SEQ ID NO:1.

The present invention includes variants of the hereinabove described polynucleotides which encode fragments, analogs and derivatives of the polypeptide characterized by the deduced amino acid sequence selected from the group consisting of the amino acid sequence set forth in SEQ ID NO:2 and SEQ ID NO:5.

The present invention also relates to vectors which include polynucleotides of the present invention, host cells which are genetically engineered with vectors of the invention and the production of polypeptides of the invention by recombinant techniques.

Also provided is an antibody against a polypeptide selected from the group consisting of the amino acid sequence set forth in SEQ ID NO:2 and SEQ ID NO:5. Still further provided is an antagonist which inhibits the activity of the polypeptide selected from the group consisting of the amino acid sequence set forth in SEQ ID NO:2 and SEQ ID NO:5.

A method is also provided for the treatment of an individual having need to inhibit novel NAGPU polypeptide comprising: administering to the individual a therapeutically effective amount of an antagonist against the polypeptide of the invention.

Provided is a process for diagnosing a disease related to expression of the polypeptide of the invention comprising: determining a nucleic acid sequence encoding the polypeptide selected from the group consisting of the amino acid sequence set forth in SEQ ID NO:2 and SEQ ID NO:5.

A diagnostic process is provided comprising: analyzing for the presence of the polypeptide selected from the group consisting of the amino acid sequence set forth in SEQ ID NO:2 and SEQ ID NO:5, either in a sample derived from a host.

In accordance with yet a further aspect of the present invention, there is provided the use of a polypeptide of the invention for therapeutic or prophylactic purposes, for example, as an antibacterial agent or a vaccine.

In accordance with another aspect of the present invention, there is provided the use of a polynucleotide of the invention for therapeutic or prophylactic purposes, in particular genetic immunization.

In accordance with yet another aspect of the present invention, there are provided inhibitors to such polypeptides, useful as antibacterial agents.

Another aspect of the invention is a pharmaceutical composition comprising the above polypeptide, polynucleotide or inhibitor of the invention and a pharmaceutically acceptable carrier.

In a particular aspect the invention provides the use of the polypeptide, polynucleotide or inhibitor of the invention to interfere with the immediate physical interaction between a pathogen and mammalian host responsible for sequelae of infection.

The invention further relates to the manufacture of a medicament for such uses.

This invention provides a method of screening drugs to identify those which interfere with the interaction of the protein or active fragment to mammalian cells.

Further provided is a method for identifying compounds which bind to and inhibit an activity of the polypeptide selected from the group consisting of the amino acid sequence set forth in SEQ ID NO:2 and SEQ ID NO:5, the method comprising: contacting a cell expressing on the surface thereof a binding means for the polypeptide, said binding means being associated with a second component capable of providing a detectable signal in response to the binding of a compound to said binding means, with a compound to be screened under conditions to permit binding to the binding means; and determining whether the compound binds to and activates or inhibits the binding by detecting the presence or absence of a signal generated from the interaction of the compound with the binding means.

BRIEF DESCRIPTION OF THE DRAWINGS

The following drawings depict certain embodiments of the invention. They are illustrative only and do not limit the invention otherwise disclosed herein.

Figure 1 (A-C) shows the polynucleotide sequence of novel *S. aureus* NAGPU gene [SEQ ID NO:1] in the antisense orientation. The complementary sequences of the two start codons are shown in bold and underlined and the complementary sequence of the stop codon is underlined. Figure 1 A-C inclusive is referred to herein as "Figure 1"

Figure 2 shows a polypeptide sequence of a novel NAGPU protein [SEQ ID NO:2] deduced from the polynucleotide sequence of Figure 1 [SEQ ID NO:1].

Figure 3 shows the oligonucleotide primers [SEQ ID NO: 3 and 4] derived from the polynucleotide sequence of Figure 1[SEQ ID NO:1].

Figure 4 shows a polypeptide sequence of a novel NAGPU protein [SEQ ID NO:5] having an alternate amino terminus deduced from the polynucleotide sequence of Figure 1 [SEQ ID NO:1].

DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to a novel N-acetylglucosamine-1-phosphate uridyltransferase protein from *Staphylococcus aureus* WCUH29, characterized in that it comprises an amino acid sequence selected from the group consisting of that given in SEQ ID NO:2 and SEQ ID NO:5, or a fragment, analogue or derivative thereof either. The amino acid sequences of SEQ ID NO: 2 and 5 are two possible translated open reading frame sequences of SEQ ID NO:1 and display homology of about 55% identity to the *B. subtilis* NAGPU enzyme sequence. Two preferred NAGPU enzymes of the invention differ at their amino terminus, each begining with a different terminal metionine. SEQ ID NO:2 and 5 show these two forms.

S. aureus WCUH 29 has been deposited at the National Collection of Industrial and Marine Bacteria Ltd. (NCIMB), Aberdeen, Scotland under number NCIMB 40771 on 11 September 1995.

The invention also relates to a polypeptide fragment of the NAGPU regulator protein, having the amino acid sequence selected from the group consisting of SEQ ID NO:2 and 5, or a derivative thereof either.

Hereinafter the term polypeptide(s) will be used to refer to the NAGPU regulator protein, its fragments, analogues or derivatives as well as the polypeptide fragment or its derivatives.

In accordance with another aspect of the present invention, there are provided polynucleotides (DNA or RNA) which encode such polypeptides.

In particular the invention provides a polynucleotide having the DNA sequence given in SEQ ID NO:1. The invention further provides a polynucleotide encoding a NAGPU protein from *S. aureus* WCUH 29 and characterized in that it comprises the DNA sequence given in SEQ ID NO:1.

The present invention also provides a novel protein from *Staphylococcus. aureus* WCUH29 obtainable by expression of a gene characterized in that it comprises the DNA sequence given SEQ ID NO:1, or a fragment, analogue or derivative thereof.

The invention also relates to novel oligonucleotides, including, for example, SEQ ID NO: 3 and 4, derived from the sequence SEQ ID NO:1 which can act as PCR primers in the process herein described to determine whether or not the *Staphylococcus aureus* genes identified herein in whole or in part are transcribed in infected tissue. The invention also relates to novel oligonucleotides, for example, SEQ ID NO: 6, that can act as hybridizations probes. It is recognized that such sequences will also have utility in diagnosis of the stage of infection and type of infection the pathogen has attained.

The polynucleotide having the DNA sequence given in SEQ ID NO:1 was obtained from the sequencing of a library of clones of chromosomal DNA of *S.aureus* WCUH 29 in *E.coli*. It has been demonstrated by the process herein described that it is transcribed *in vivo* in an established infection of *S.aureus* WCUH29 in a mouse model of infection.

To obtain the polynucleotide encoding the protein using the DNA sequence given in SEQ ID NO:1 typically a library of clones of chromosomal DNA of *S.aureus* WCUH 29 in *E.coli* or some other suitable host is probed with a radiolabelled oligonucleotide, preferably a 17-mer or longer, derived from the partial sequence. Clones carrying DNA identical to that of the probe can then be distinguished using high stringency washes. By sequencing the individual clones thus identified with sequencing primers designed from the original sequence it is then possible to extend the sequence in both directions to determine the full gene sequence. Conveniently such sequencing is performed using denatured double

stranded DNA prepared from a plasmid clone. Suitable techniques are described by Maniatis, T., Fritsch, E.F. and Sambrook, J. in MOLECULAR CLONING, A Laboratory Manual [2nd edition 1989 Cold Spring Harbor Laboratory, see Screening By Hybridization 1.90 and Sequencing Denatured Double-Stranded DNA Templates 13.70].

The polynucleotide of the present invention may be in the form of RNA or in the form of DNA, which DNA includes cDNA, genomic DNA, and synthetic DNA. The DNA may be double-stranded or single-stranded, and if single stranded may be the coding strand or non-coding (anti-sense) strand. The coding sequence which encodes the polypeptide may be identical to the coding sequence shown in SEQ ID NO:1 or may be a different coding sequence which coding sequence, as a result of the redundancy or degeneracy of the genetic code, encodes the same polypeptide.

The present invention includes variants of the hereinabove described polynucleotides which encode fragments, analogues and derivatives of the polypeptide characterized by the deduced amino acid sequence selected from the group consisting of SEQ ID NO:2 and/or 5. The variant of the polynucleotide may be a naturally occurring allelic variant of the polynucleotide or a non-naturally occurring variant of either polynucleotide.

Thus, the present invention includes polynucleotides encoding the same polypeptide characterized by the deduced amino acid sequence selected from the group consisting of SEQ ID NO:2 and/or 5, as well as variants of such polynucleotides which variants encode for a fragment, derivative or analogue of either polypeptide. Such nucleotide variants include deletion variants, substitution variants and addition or insertion variants.

The polynucleotide may have a coding sequence which is a naturally occurring allelic variant of the coding sequence characterized by the DNA sequence of SEQ ID NO:1. As known in the art, an allelic variant is an alternate form of a polynucleotide sequence which may have a substitution, deletion or addition of one or more nucleotides, which does not substantially alter the function of the encoded polypeptide.

The polynucleotide which encodes for the mature polypeptide, i.e. the NAGPU protein, may include only the coding sequence for the mature polypeptide or the coding sequence for the mature polypeptide and additional coding sequence such as a leader or secretory sequence or a proprotein sequence.

Thus, the term "polynucleotide encoding a polypeptide" encompasses a polynucleotide which includes only coding sequence for the polypeptide as well as a polynucleotide which includes additional coding and/or non-coding sequence.

The present invention therefore includes polynucleotides, wherein the coding sequence for the mature polypeptide may be fused in the same reading frame to a

polynucleotide sequence which aids in expression and secretion of a polypeptide from a host cell, for example, a leader sequence which functions as a secretory sequence for controlling transport of a polypeptide from the cell. The polypeptide having a leader sequence is a preprotein and may have the leader sequence cleaved by the host cell to form the mature form of the polypeptide. The polynucleotides may also encode for a proprotein which is the mature protein plus additional 5' amino acid residues. A mature protein having a prosequence is a proprotein and is an inactive form of the protein. Once the prosequence is cleaved an active mature protein remains.

Thus, for example, the polynucleotide of the present invention may encode for a mature protein, or for a protein having a prosequence or for a protein having both a prosequence and a presequence (leader sequence). During post-translational modification of the peptide, a methionine residue at the NH_2-terminus may be deleted. Accordingly, this invention contemplates the use of both the methionine-containing and the methionineless amino terminal variants of the protein of the invention.

The polynucleotides of the present invention may also have the coding sequence fused in frame to a marker sequence at either the 5' or 3' terminus of the gene which allows for purification of the polypeptide of the present invention. The marker sequence may be a hexa-histidine tag supplied by the pQE series of vectors (supplied commercially by Quiagen Inc.) to provide for purification of the polypeptide fused to the marker in the case of a bacterial host.

The present invention further relates to polynucleotides which hybridize to the hereinabove-described sequences if there is at least 50% and preferably at least 70% identity between the sequences. The present invention particularly relates to polynucleotides which hybridize under stringent conditions to the hereinabove-described polynucleotides . As herein used, the term "stringent conditions" means hybridization will occur only if there is at least 95% and preferably at least 97% identity between the sequences. The polynucleotides which hybridize to the hereinabove described polynucleotides in a preferred embodiment encode polypeptides which retain substantially the same biological function or activity as the polypeptide characterized by the deduced amino acid sequence selected from the group consisting of SEQ ID NO:2 and/or 5.

The deposit referred to herein will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Micro-organisms for purposes of Patent Procedure. These deposits are provided merely as convenience to those of skill in the art and are not an admission that a deposit is required under 35 U.S.C. §112. The sequence of the polynucleotides contained in the deposited material, as well as the amino acid sequence of the polypeptides encoded thereby, are incorporated herein by reference and are controlling in the event of any conflict with any description of sequences herein. A

license may be required to make, use or sell the deposited material, and no such license is hereby granted.

The terms "fragment," "derivative" and "analogue" when referring to the polypeptide characterized by the deduced amino acid sequence selected from the group consisting of SEQ ID NO:2 and/or 5, means a polypeptide which retains essentially the same biological function or activity as such polypeptide. Thus, an analogue includes a proprotein which can be activated by cleavage of the proprotein portion to produce an active mature polypeptide.

The polypeptide of the present invention may be a recombinant polypeptide, a natural polypeptide or a synthetic polypeptide, preferably a recombinant polypeptide.

The fragment, derivative or analogue of the polypeptide characterized by the deduced amino acid sequence selected from the group consisting of SEQ ID NO:2 and/or 5 may be (i) one in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue) and such substituted amino acid residue may or may not be one encoded by the genetic code, or (ii) one in which one or more of the amino acid residues includes a substituent group, or (iii) one in which the polypeptide is fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol), or (iv) one in which the additional amino acids are fused to the polypeptide, such as a leader or secretory sequence or a sequence which is employed for purification of the polypeptide or a proprotein sequence. Such fragments, derivatives and analogues are deemed to be within the scope of those skilled in the art from the teachings herein.

The polypeptides and polynucleotides of the present invention are preferably provided in an isolated form, and preferably are purified to homogeneity.

The term "isolated" means that the material is removed from its original environment (e.g., the natural environment if it is naturally occurring). For example, a naturally-occurring polynucleotide or polypeptide present in a living animal is not isolated, but the same polynucleotide or polypeptide, separated from some or all of the coexisting materials in the natural system, is isolated. Such polynucleotides could be part of a vector and/or such polynucleotides or polypeptides could be part of a composition, and still be isolated in that such vector or composition is not part of its natural environment.

The present invention also relates to vectors which include polynucleotides of the present invention, host cells which are genetically engineered with vectors of the invention and the production of polypeptides of the invention by recombinant techniques.

In accordance with yet a further aspect of the present invention, there is therefore provided a process for producing the polypeptide of the invention by recombinant techniques by expressing a polynucleotide encoding said polypeptide in a host and

recovering the expressed product. Alternatively, the polypeptides of the invention can be synthetically produced by conventional peptide synthesizers.

Host cells are genetically engineered (transduced or transformed or transfected) with the vectors of this invention which may be, for example, a cloning vector or an expression vector. The vector may be, for example, in the form of a plasmid, a cosmid, a phage, etc. The engineered host cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants or amplifying the genes. The culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and will be apparent to the ordinarily skilled artisan.

Suitable expression vectors include chromosomal, nonchromosomal and synthetic DNA sequences, e.g., bacterial plasmids; phage DNA; baculovirus; yeast plasmids; vectors derived from combinations of plasmids and phage DNA. However, any other vector may be used as long as it is replicable and viable in the host.

The appropriate DNA sequence may be inserted into the vector by a variety of procedures. In general, the DNA sequence is inserted into an appropriate restriction endonuclease site(s) by procedures known in the art.

The DNA sequence in the expression vector is operatively linked to an appropriate expression control sequence(s) (promoter) to direct mRNA synthesis. As representative examples of such promoters, there may be mentioned: LTR or SV40 promoter, the *E. coli. lac* or *trp*, the phage lambda P_L promoter and other promoters known to control expression of genes in eukaryotic or prokaryotic cells or their viruses. The expression vector also contains a ribosome binding site for translation initiation and a transcription terminator. The vector may also include appropriate sequences for amplifying expression.

In addition, the expression vectors preferably contain one or more selectable marker genes to provide a phenotypic trait for selection of transformed host cells such as dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, or such as tetracycline or ampicillin resistance in *E. coli*.

The gene can be placed under the control of a promoter, ribosome binding site (for bacterial expression) and, optionally, an operator (collectively referred to herein as "control" elements), so that the DNA sequence encoding the desired protein is transcribed into RNA in the host cell transformed by a vector containing this expression construction. The coding sequence may or may not contain a signal peptide or leader sequence. The polypeptides of the present invention can be expressed using, for example, the *E. coli* tac promoter or the protein A gene (spa) promoter and signal sequence. Leader sequences can be removed by the bacterial host in post-translational processing. See, e.g., U.S. Patent Nos. 4,431,739; 4,425,437; 4,338,397. Promoter regions can be selected from any desired

gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are PKK232-8 and PCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda P_R, P_L and trp. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art.

In addition to control sequences, it may be desirable to add regulatory sequences which allow for regulation of the expression of the protein sequences relative to the growth of the host cell. Regulatory sequences are known to those of skill in the art, and examples include those which cause the expression of a gene to be turned on or off in response to a chemical or physical stimulus, including the presence of a regulatory compound. Other types of regulatory elements may also be present in the vector, for example, enhancer sequences.

An expression vector is constructed so that the particular coding sequence is located in the vector with the appropriate regulatory sequences, the positioning and orientation of the coding sequence with respect to the control sequences being such that the coding sequence is transcribed under the "control" of the control sequences (i.e., RNA polymerase which binds to the DNA molecule at the control sequences transcribes the coding sequence). Modification of the coding sequences may be desirable to achieve this end. For example, in some cases it may be necessary to modify the sequence so that it may be attached to the control sequences with the appropriate orientation; i.e., to maintain the reading frame. The control sequences and other regulatory sequences may be ligated to the coding sequence prior to insertion into a vector, such as the cloning vectors described above. Alternatively, the coding sequence can be cloned directly into an expression vector which already contains the control sequences and an appropriate restriction site.

Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, e.g., the ampicillin resistance gene of *E. coli* and *S. cerevisiae* TRP1 gene, and a promoter derived from a highly-expressed gene to direct transcription of a downstream structural sequence. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably, a leader sequence capable of directing secretion of translated protein into the periplasmic space or extracellular medium. Optionally, the heterologous sequence can encode a fusion protein including an N-terminal identification peptide imparting desired characteristics, e.g., stabilization or simplified purification of expressed recombinant product.

The vector containing the appropriate DNA sequence as hereinabove described, as well as an appropriate promoter or control sequence, may be employed to transform an appropriate host to permit the host to express the protein.

More particularly, the present invention also includes recombinant constructs comprising one or more of the sequences as broadly described above. The constructs comprise a vector, such as a plasmid or viral vector, into which a sequence of the invention has been inserted, in a forward or reverse orientation. In a preferred aspect of this embodiment, the construct further comprises regulatory sequences, including, for example, a promoter, operably linked to the sequence. Large numbers of suitable vectors and promoters are known to those of skill in the art, and are commercially available. The following vectors are provided by way of example. Bacterial: pET-3 vectors (Stratagene), pQE70, pQE60, pQE-9 (Qiagen), pbs, pD10, phagescript, psiX174, pbluescript SK, pbsks, pNH8A, pNH16a, pNH18A, pNH46A (Stratagene); ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia). Eukaryotic: pBlueBacIII (Invitrogen), pWLNEO, pSV2CAT, pOG44, pXT1, pSG (Stratagene) pSVK3, pBPV, pMSG, pSVL (Pharmacia). However, any other plasmid or vector may be used as long as they are replicable and viable in the host.

Examples of recombinant DNA vectors for cloning and host cells which they can transform include the bacteriophage λ (*E. coli*), pBR322 (*E. coli*), pACYC177 (*E. coli*), pKT230 (gram-negative bacteria), pGV1106 (gram-negative bacteria), pLAFR1 (gram-negative bacteria), pME290 (non-*E. coli* gram-negative bacteria), pHV14 (*E. coli* and *Bacillus subtilis*), pBD9 (*Bacillus*), pIJ61 (*Streptomyces*), pUC6 (*Streptomyces*), YIp5 (*Saccharomyces*), a baculovirus insect cell system, , YCp19 (*Saccharomyces*). See, generally, "DNA Cloning": Vols. I & II, Glover *et al.* ed. IRL Press Oxford (1985) (1987) and: T. Maniatis *et al.* ("Molecular Cloning" Cold Spring Harbor Laboratory (1982).

In some cases, it may be desirable to add sequences which cause the secretion of the polypeptide from the host organism, with subsequent cleavage of the secretory signal.

Polypeptides can be expressed in host cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook, et al., Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor, N.Y., (1989), the disclosure of which is hereby incorporated by reference.

Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter is induced by appropriate means (e.g., temperature shift or chemical induction) and cells are cultured for an additional period.

Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents, such methods are well known to those skilled in the art.

Depending on the expression system and host selected, the polypeptide of the present invention may be produced by growing host cells transformed by an expression vector described above under conditions whereby the polypeptide of interest is expressed. The polypeptide is then isolated from the host cells and purified. If the expression system secretes the polypeptide into growth media, the polypeptide can be purified directly from the media. If the polypeptide is not secreted, it is isolated from cell lysates or recovered from the cell membrane fraction. Where the polypeptide is localized to the cell surface, whole cells or isolated membranes can be used as an assayable source of the desired gene product. Polypeptide expressed in bacterial hosts such as *E. coli* may require isolation from inclusion bodies and refolding. Where the mature protein has a very hydrophobic region which leads to an insoluble product of overexpression, it may be desirable to express a truncated protein in which the hydrophobic region has been deleted. The selection of the appropriate growth conditions and recovery methods are within the skill of the art.

The polypeptide can be recovered and purified from recombinant cell cultures by methods including ammonium sulphate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps.

Depending upon the host employed in a recombinant production procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated. Polypeptides of the invention may also include an initial methionine amino acid residue.

A "replicon" is any genetic element (e.g., plasmid, chromosome, virus) that functions as an autonomous unit of DNA replication *in vivo*; i.e., capable of replication under its own control.

A "vector" is a replicon, such as a plasmid, phage, or cosmid, to which another DNA segment may be attached so as to bring about the replication of the attached segment.

A "double-stranded DNA molecule" refers to the polymeric form of deoxyribonucleotides (bases adenine, guanine, thymine, or cytosine) in a double-stranded helix, both relaxed and supercoiled. This term refers only to the primary and secondary structure of the molecule, and does not limit it to any particular tertiary forms. Thus, this

term includes double-stranded DNA found, *inter alia*, in linear DNA molecules (e.g., restriction fragments), viruses, plasmids, and chromosomes. In discussing the structure of particular double-stranded DNA molecules, sequences may be described herein according to the normal convention of giving only the sequence in the 5' to 3' direction along the nontranscribed strand of DNA (i.e., the strand having the sequence homologous to the mRNA).

A DNA "coding sequence of" or a "nucleotide sequence encoding" a particular protein, is a DNA sequence which is transcribed and translated into a polypeptide when placed under the control of appropriate regulatory sequences.

A "promoter sequence" is a DNA regulatory region capable of binding RNA polymerase in a cell and initiating transcription of a downstream (3' direction) coding sequence. For purposes of defining the present invention, the promoter sequence is bound at the 3' terminus by a translation start codon (e.g., ATG) of a coding sequence and extends upstream (5' direction) to include the minimum number of bases or elements necessary to initiate transcription at levels detectable above background. Within the promoter sequence will be found a transcription initiation site (conveniently defined by mapping with nuclease S1), as well as protein binding domains (consensus sequences) responsible for the binding of RNA polymerase. Eukaryotic promoters will often, but not always, contain "TATA" boxes and "CAT" boxes. Prokaryotic promoters contain Shine-Dalgarno sequences in addition to the -10 and -35 consensus sequences.

DNA "control sequences" refers collectively to promoter sequences, ribosome binding sites, polyadenylation signals, transcription termination sequences, upstream regulatory domains, enhancers, and the like, which collectively provide for the expression (i.e., the transcription and translation) of a coding sequence in a host cell.

A control sequence "directs the expression" of a coding sequence in a cell when RNA polymerase will bind the promoter sequence and transcribe the coding sequence into mRNA, which is then translated into the polypeptide encoded by the coding sequence.

A "host cell" is a cell which has been transformed or transfected, or is capable of transformation or transfection by an exogenous DNA sequence.

A cell has been "transformed" by exogenous DNA when such exogenous DNA has been introduced inside the cell membrane. Exogenous DNA may or may not be integrated (covalently linked) into chromosomal DNA making up the genome of the cell. In prokaryotes and yeasts, for example, the exogenous DNA may be maintained on an episomal element, such as a plasmid. With respect to eukaryotic cells, a stably transformed or transfected cell is one in which the exogenous DNA has become integrated into the chromosome so that it is inherited by daughter cells through chromosome replication. This

stability is demonstrated by the ability of the eukaryotic cell to establish cell lines or clones comprised of a population of daughter cells containing the exogenous DNA.

A "clone" is a population of cells derived from a single cell or common ancestor by mitosis. A "cell line" is a clone of a primary cell that is capable of stable growth *in vitro* for many generations.

A "heterologous" region of a DNA construct is an identifiable segment of DNA within or attached to another DNA molecule that is not found in association with the other molecule in nature.

Each of the DNA sequences provided herein may be used in the discovery and development of antibacterial compounds. The encoded protein upon expression can be used as a target for the screening of antibacterial drugs. Additionally, the DNA sequences encoding the amino terminal regions of the encoded protein or Shine-Delgarno or other translation facilitating sequences of the respective mRNA can be used to construct antisense sequences to control the expression of the coding sequence of interest.

In accordance with yet another aspect of the present invention, there are provided inhibitors to such polypeptides, useful as antibacterial agents. Inhibitors of NAGPU proteins have utility in anti-bacterial therapy.

Another aspect of the invention is a pharmaceutical composition comprising an inhibitor of the invention and a pharmaceutically acceptable carrier.

In a particular aspect the invention provides the use of an inhibitor of the invention as an antibacterial agent.

The invention further relates to the manufacture of a medicament for such uses.

This invention provides a method of screening drugs to identify those which are antibacterial by measuring the ability of the drug to interfere with the biosynthesis of uridyl diphosphate N-acetyl glucosamine by the NAGPU protein.

It has been shown that *E.coli* NAGPU will act as a pyrophosphorylase, catalyzing the reverse reaction to N-acetyl glucosamine-1-phosphate from the products of the forward reaction, UDP-N-acetyl glucosamine and pyrophosphate (Strominger, J.R. and Smith, M.S. [1959] J. Biol. Chem. 234: 1822-7). By introducing an inorganic pyrophosphatase into the reaction it will proceed in the forward direction without limit (Mengin-Lecreulx, D. and van Heijenoort, J., J. Bacteriol. 176: 5788-5795 [1994]).

In a preferred embodiment, N-acetylglucosamine-1-phosphate is incubated with UTP and inorganic pyrophosphatase in the presence of the NAGPU protein to generate inorganic phosphate which can be measured colorimetrically using a suitably sensitive procedure such as the Malachite Green method (Itaya,K. & Ui,M. Clin.Chim.Acta 14,361-366 [1966) to provide a measurement of NAGPU enzymatic activity. The decrease of enzymatic activity in this reaction would indicate the presence of an inhibitor.

The invention also relates to inhibitors identified thereby.

In therapy or as a prophylactic, the active agent may be administered to a patient as an injectable composition, for example as a sterile aqueous dispersion, preferably isotonic.

Alternatively the composition may be formulated for topical application for example in the form of ointments, creams, lotions, eye ointments, eye drops, ear drops, mouthwash, impregnated dressings and sutures and aerosols, and may contain appropriate conventional additives, including, for example, preservatives, solvents to assist drug penetration, and emollients in ointments and creams. Such topical formulations may also contain compatible conventional carriers, for example cream or ointment bases, and ethanol or oleyl alcohol for lotions. Such carriers may constitute from about 1% to about 98% by weight of the formulation; more usually they will constitute up to about 80% by weight of the formulation.

For administration to human patients, it is expected that the daily dosage level of the active agent will be from 0.01 to 10 mg/kg, typically around 1 mg/kg. The physician in any event will determine the actual dosage which will be most suitable for an individual patient and will vary with the age, weight and response of the particular patient. The above dosages are exemplary of the average case. There can, of course, be individual instances where higher or lower dosage ranges are merited, and such are within the scope of this invention.

Within the indicated dosage range, no adverse toxicological effects are expected with the compounds of the invention which would preclude their administration to suitable patients.

Examples

The present invention is further described by the following examples. These exemplification's, while illustrating certain specific aspects of the invention, do not portray the limitations or circumscribe the scope of the disclosed invention.

In order to facilitate understanding of the following examples certain frequently occurring methods and/or terms will be described.

"Plasmids" are designated by a lower case p preceded and/or followed by capital letters and/or numbers. The starting plasmids herein are either commercially available, publicly available on an unrestricted basis, or can be constructed from available plasmids in accord with published procedures. In addition, equivalent plasmids to those described are known in the art and will be apparent to the ordinarily skilled artisan.

"Digestion" of DNA refers to catalytic cleavage of the DNA with a restriction enzyme that acts only at certain sequences in the DNA. The various restriction enzymes used herein are commercially available and their reaction conditions, cofactors and other

requirements were used as would be known to the ordinarily skilled artisan. For analytical purposes, typically 1 μg of plasmid or DNA fragment is used with about 2 units of enzyme in about 20 μl of buffer solution. For the purpose of isolating DNA fragments for plasmid construction, typically 5 to 50 μg of DNA are digested with 20 to 250 units of enzyme in a larger volume. Appropriate buffers and substrate amounts for particular restriction enzymes are specified by the manufacturer. Incubation times of about 1 hour at 37° C are ordinarily used, but may vary in accordance with the supplier's instructions. After digestion the reaction is electrophoresed directly on a polyacrylamide gel to isolate the desired fragment.

Size separation of the cleaved fragments is performed using 8 percent polyacrylamide gel described by Goeddel, D. *et al.*, Nucleic Acids Res., 8:4057 (1980).

"Oligonucleotides" refers to either a single stranded polydeoxynucleotide or two complementary polydeoxynucleotide strands which may be chemically synthesized. Such synthetic oligonucleotides have no 5' phosphate and thus will not ligate to another oligonucleotide without adding a phosphate with an ATP in the presence of a kinase. A synthetic oligonucleotide will ligate to a fragment that has not been dephosphorylated.

"Ligation" refers to the process of forming phosphodiester bonds between two double stranded nucleic acid fragments (Maniatis, T., et al., Id., p. 146). Unless otherwise provided, ligation may be accomplished using known buffers and conditions with 10 units to T4 DNA ligase ("ligase") per 0.5 μg of approximately equimolar amounts of the DNA fragments to be ligated.

All examples were carried out using standard techniques, which are well known and routine to those of skill in the art, except where otherwise described in detail. Routine molecular biology techniques of the following examples can be carried out as described in standard laboratory manuals, such as Sambrook et al., *MOLECULAR CLONING: A LABORATORY MANUAL*, 2nd Ed.; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989).

Example 1

Isolation of DNA coding for a NAGPU Protein from *S. Aureus* WCUH 29

The polynucleotide having the DNA sequence given in SEQ ID NO:1 was obtained from a library of clones of chromosomal DNA of *S.aureus* WCUH 29 in *E.coli*. In some cases the sequencing data from two or more clones containing overlapping *S.aureus* WCUH 29 DNA was used to construct the contiguous DNA sequence in SEQ ID NO:1.

Libraries may be prepared by routine methods, for example:

Methods 1 and 2

Total cellular DNA is isolated from *Staphylococcus aureus* strain WCUH29 (NCIMB 40771) according to standard procedures and size-fractionated by either of two methods.

Method 1.

Total cellular DNA is mechanically sheared by passage through a needle in order to size-fractionate according to standard procedures. DNA fragments of up to 11kbp in size are rendered blunt by treatment with exonuclease and DNA polymerase, and EcoRI linkers added. Fragments are ligated into the vector Lambda ZapII that has been cut with EcoRI, the library packaged by standard procedures and *E.coli* infected with the packaged library. The library is amplified by standard procedures.

Method 2.

Total cellular DNA is partially hydrolyzed with a combination of four restriction enzymes (RsaI, PalI, AluI and Bsh1235I) and size-fractionated according to standard procedures. EcoRI linkers are ligated to the DNA and the fragments then ligated into the vector Lambda ZapII that have been cut with EcoRI, the library packaged by standard procedures, and *E.coli* infected with the packaged library. The library is amplified by standard procedures.

Example 2

The determination of expression during infection of a gene from *Staphylococcus aureus* WCUH29

Necrotic fatty tissue from a four day groin infection of *Staphylococcus aureus* WCUH29 in the mouse is efficiently disrupted and processed in the presence of chaotropic agents and RNAase inhibitor to provide a mixture of animal and bacterial RNA. The optimal conditions for disruption and processing to give stable preparations and high yields of bacterial RNA are followed by the use of hybridization to a radiolabelled oligonucleotide specific to *Staphylococcus aureus* 16S RNA on Northern blots. The RNAase free, DNAase free, DNA and protein free preparations of RNA obtained are suitable for Reverse Transcription PCR (RT-PCR) using unique primer pairs designed from the sequence of each gene of *Staphylococcus aureus* WCUH29.

a) Isolation of tissue infected with Staphylococcus aureus WCUH29 from a mouse animal model of infection

10 ml. volumes of sterile nutrient broth (No.2 Oxoid) are seeded with isolated, individual colonies of Staphylococcus aureus WCUH29 from an agar culture plate. The cultures are incubated aerobically (static culture) at 37 degrees C for 16-20 hours . 4 week old mice (female,18g-22g, strain MF1) are each infected by subcutaneous injection of 0.5ml. of this broth culture of Staphylococcus aureus WCUH29 (diluted in broth to approximately 10^8 cfu/ml.) into the anterior , right lower quadrant (groin area). Mice should

be monitored regularly during the first 24 hours after infection, then daily until termination of study. Animals with signs of systemic infection, i.e. lethargy, ruffled appearance, isolation from group, should be monitored closely and if signs progress to moribundancy, the animal should be culled immediately.

Visible external signs of lesion development will be seen 24-48h after infection. Examination of the abdomen of the animal will show the raised outline of the abscess beneath the skin. The localized lesion should remain in the right lower quadrant, but may occasionally spread to the left lower quadrant, and superiorly to the thorax. On occasions, the abscess may rupture through the overlying skin layers. In such cases the affected animal should be culled immediately and the tissues sampled if possible. Failure to cull the animal may result in the necrotic skin tissue overlying the abscess being sloughed off, exposing the abdominal muscle wall.

Approximately 96h after infection, animals are killed using carbon dioxide asphyxiation. To minimize delay between death and tissue processing /storage, mice should be killed individually rather than in groups. The dead animal is placed onto its back and the fur swabbed liberally with 70% alcohol. An initial incision using scissors is made through the skin of the abdominal left lower quadrant, traveling superiorly up to, then across the thorax. The incision is completed by cutting inferiorly to the abdominal lower right quadrant. Care should be taken not to penetrate the abdominal wall. Holding the skin flap with forceps, the skin is gently pulled way from the abdomen. The exposed abscess, which covers the peritoneal wall but generally does not penetrate the muscle sheet completely, is excised, taking care not to puncture the viscera

The abscess/muscle sheet and other infected tissue may require cutting in sections, prior to flash-freezing in liquid nitrogen, thereby allowing easier storage in plastic collecting vials.

b) Isolation of *Staphylococcus aureus* WCUH29 RNA from infected tissue samples

4-6 infected tissue samples(each approximately 0.5-0.7g) in 2ml screw-cap tubes are removed from -80°C storage into a dry ice ethanol bath In a microbiological safety cabinet the samples are disrupted individually whilst the remaining samples are kept cold in the dry ice ethanol bath. To disrupt the bacteria within the tissue sample 1ml of TRIzol Reagent (Gibco BRL, Life Technologies) is added followed by enough 0.1mm zirconia/silica beads to almost fill the tube, the lid is replaced taking care not to get any beads into the screw thread so as to ensure a good seal and eliminate aerosol generation. The sample is then homogenized in a Mini-BeadBeater Type BX-4 (Biospec Products). Necrotic fatty tissue is treated for 100 seconds at 5000 rpm in order to achieve bacterial lysis. *In vivo* grown bacteria require longer treatment than *in vitro* grown *S.aureus* WCUH29 which are disrupted by a 30 second bead-beat.

After bead-beating the tubes are chilled on ice before opening in a fume-hood as heat generated during disruption may degrade the TRIzol and release cyanide.

200 microlitres of chloroform is then added and the tubes shaken by hand for 15 seconds to ensure complete mixing. After 2-3 minutes at room temperature the tubes are spun down at 12,000 x g, 4 °C for 15 minutes and RNA extraction is then continued according to the method given by the manufacturers of TRIzol Reagent i.e.:- The aqueous phase, approximately 0.6 ml, is transferred to a sterile Eppendorf tube and 0.5 ml of isopropanol is added. After 10 minutes at room temperature the samples are spun at 12,000 x g, 4 °C for 10 minutes. The supernatant is removed and discarded then the RNA pellet is washed with 1 ml 75% ethanol. A brief vortex is used to mix the sample before centrifuging at 7,500 x g, 4 °C for 5 minutes. The ethanol is removed and the RNA pellet dried under vacuum for no more than 5 minutes. Samples are then resuspended by repeated pipetting in 100 microlitres of DEPC treated water, followed by 5-10 minutes at 55 °C. Finally, after at least 1 minute on ice, 200 units of Rnasin (Promega) is added.

RNA preparations are stored at -80 °C for up to one month. For longer term storage the RNA precipitate can be stored at the wash stage of the protocol in 75% ethanol for at least one year at -20 °C.

Quality of the RNA isolated is assessed by running samples on 1% agarose gels. 1 x TBE gels stained with ethidium bromide are used to visualize total RNA yields. To demonstrate the isolation of bacterial RNA from the infected tissue 1 x MOPS, 2.2M formaldehyde gels are run and vacuum blotted to Hybond-N (Amersham). The blot is then hybridized with a ^{32}P labeled oligonucletide probe specific to 16s rRNA of *S.aureus* (K.Greisen, M. Loeffelholz, A. Purohit and D. Leong. J.Clin. (1994) Microbiol. 32 335-351). An oligonucleotide of the sequence: 5'-gctcctaaaaggttactccaccggc-3' [SEQ ID NO:6] is used as a probe. The size of the hybridizing band is compared to that of control RNA isolated from *in vitro* grown *S.aureus* WCUH29 in the Northern blot. Correct sized bacterial 16s rRNA bands can be detected in total RNA samples which show extensive degradation of the mammalian RNA when visualized on TBE gels.

c) The removal of DNA from Staphylococcus aureus WCUH29 derived RNA

DNA was removed from 73 microlitre samples of RNA by a 15 minute treatment on ice with 3 units of DNAasel, amplification grade (Gibco BRL, Life Technologies) in the buffer supplied with the addition of 200 units of Rnasin (Promega) in a final volume of 90 microlitres.

The DNAase was inactivated and removed by treatment with TRIzol LS Reagent (Gibco BRL, Life Technologies) according to the manufacturers protocol.
DNAase treated RNA was resuspended in 73 microlitres of DEPC treated water with the addition of Rnasin as described in Method 1.

d) The preparation of cDNA from RNA samples derived from infected tissue

10 microlitre samples of DNAase treated RNA are reverse transcribed using a SuperScript Preamplification System for First Strand cDNA Synthesis kit (Gibco BRL, Life Technologies) according to the manufacturers instructions. 1 nanogram of random hexamers is used to prime each reaction. Controls without the addition of SuperScriptII reverse transcriptase are also run. Both +/-RT samples are treated with RNaseH before proceeding to the PCR reaction

e) The use of PCR to determine the presence of a bacterial cDNA species

PCR reactions are set up on ice in 0.2ml tubes by adding the following components:

45 microlitres PCR SUPERMIX (Gibco BRL, Life Technologies).

1 microlitre 50mM $MgCl_2$, to adjust final concentration to 2.5mM.

1 microlitre PCR primers(optimally 18-25 basepairs in length and designed to possess similar annealing temperatures), each primer at 10mM initial concentration.

2 microlitres cDNA.

PCR reactions are run on a Perkin Elmer GeneAmp PCR System 9600 as follows:

5 minutes at 95 °C, then 50 cycles of 30 seconds each at 94 °C, 42 °C and 72 °C followed by 3 minutes at 72 °C and then a hold temperature of 4 °C. (the number of cycles is optimally 30-50 to determine the appearance or lack of a PCR product and optimally 8-30 cycles if an estimation of the starting quantity of cDNA from the RT reaction is to be made).

10 microlitre aliquots are then run out on 1% 1 x TBE gels stained with ethidium bromide with PCR product, if present, sizes estimated by comparison to a 100 bp DNA Ladder (Gibco BRL, Life Technologies). Alternatively if the PCR products are conveniently labeled by the use of a labelled PCR primer (e.g. labelled at the 5'end with a dye) a suitable aliquot of the PCR product is run out on a polyacrylamide sequencing gel and its presence and quantity detected using a suitable gel scanning system (e.g. ABI Prism™ 377 Sequencer using GeneScan™ software as supplied by Perkin Elmer)

RT/PCR controls may include +/- reverse transcriptase reactions, 16s rRNA primers or DNA specific primer pairs designed to produce PCR products from non-transcribed *S.aureus* WCUH29 genomic sequences.

To test the efficiency of the primer pairs they are used in DNA PCR with WCUH29 total DNA. PCR reactions are set up and run as described above using approx. 1 microgram of DNA in place of the cDNA and 35 cycles of PCR.

Primer pairs which fail to give the predicted sized product in either DNA PCR or RT/PCR are PCR failures and as such are uninformative. Of those which give the correct size product with DNA PCR two classes are distinguished in RT/PCR:

1.Genes which are not transcribed *in vivo* reproducibly fail to give a product in RT/PCR.

2.Genes which are transcribed *in vivo* reproducibly give the correct size product in RT/PCR and show a stronger signal in the +RT samples than the signal (if at all present) in -RT controls.

The following nucleotide sequence (SEQ ID NO:1) was identified in the above test as transcribed *in vivo*. Deduced amino acid sequences are given as SEQ ID NO:2 and 5. The pair of PCR primers used to identify the gene are given as SEQ ID NOs 3 and 4.

SEQUENCE LISTING

(1) GENERAL INFORMATION

(i) APPLICANT: SmithKline Beecham p.l.c.

(ii) TITLE OF THE INVENTION: NOVEL NAPU

(iii) NUMBER OF SEQUENCES: 6

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: SmithKline Beecham Corporation
(B) STREET: 709 Swedeland Road
(C) CITY: King of Prussia
(D) STATE: PA
(E) COUNTRY: U.S.A.
(F) ZIP: 19406-0939

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Diskette
(B) COMPUTER: IBM Compatible
(C) OPERATING SYSTEM: DOS
(D) SOFTWARE: FastSEQ Version 1.5

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE: 05-DEC-1996
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: 60/027972
(B) FILING DATE: 08-OCT-1996

(A) APPLICATION NUMBER: 9604580.2
(B) FILING DATE: 04-MAR-1996

(A) APPLICATION NUMBER: 9612887.1
(B) FILING DATE: 26-MAY-1996

(A) APPLICATION NUMBER: 9604581.0
(B) FILING DATE: 04-MAR-1996

(A) APPLICATION NUMBER: 9525246.6
(B) FILING DATE: 11-DEC-1995

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Gimmi, Edward R
(B) REGISTRATION NUMBER: 38,891
(C) REFERENCE/DOCKET NUMBER: P31302

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 610-270-4478
(B) TELEFAX: 610-270-5090
(C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1449 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

AATATTACAT AGGATTATCC TTTTATCCTA GCCATTTTAA ATACGTAAAT GATTATTTCC 60
TATATCCTTC TTTTGTTGTT TGTCTTGCTC TTGCCACAGC TAAACTGTCA TTTGGTACGT 120
CATCTGTGAT TGTGGAACCA GCTGCCACCA ATACATCATC ACCAATTGTT ACAGGTGCTA 180
CTAAATTAAC ATTGCAACCT ACAAATGAAT CTTTGCCGAC GATAGTTTTA AATTTATTTT 240
CACCATCATA GTTAACTGTA ATCGTTCCGC AACCAATATT AGTACGTTCG CCAATTACAG 300
CATCGCCAAT ATAACTTAAA TGTGAAACCT TGGCACCATC TTTAAGATCT GCTTTTTTAA 360
TTTCTACAAA ATTTCCAACC TTAACATCTG CACCTAATTG CGCGCCTGGT CTCAATTGCG 420
CAAACGGTCC GACCTTAGTA TTAGCTCCTA CGCTAGCATC ATTAACAACA GACTGTTGAA 480
TACATGCACC ATTTTCAATC GTACTATTGT TAATTTCAGA GTACTGACCA ATAACAACAT 540
CTTCGCCAAT TTCTGTACGA CCATTAATTC GTACGCCTGG TTCAATGACT GTATCACTAC 600
CAATTGTAAC GTCTGGACCA ATAAAAGTGC TGTCAGGATC GATGATTGTC ACACCATTTA 660
GCATGTGATA ATGATTCGTA CGACGTTGCA TCGCCTTCTC AGCCTGACTA AGCATTACAC 720
GATCATTTAC ACCCATGATT TCTTCAACAT CATTGGTACG ATACACTTCT ACGATGCCGC 780
CATCATTTAA AATTAACGAC AATACATCAG GGAGGTAATA TTCACCTTGC GCATTATCAT 840
TTTTCACTTG TGTTAATTTT TCAAACAACG TTTTATTATT AAACGCAAAA ATACCTGAAC 900
TAATTTCATT AATATCCTTT TCAGCTTGCG TTGCATCTTT CTCTTCAACT ATGCGTTCTA 960
AACGACCTGA CGCATTTCGA ACGATTCTTC CGTATCCATA TGGTTGTTGA ATCGATGCAG 1020

ATAATACAGT TGCTTGAGCA TTAGCATCCT CATGATGCGC AATCAATGTT TCTAATGTTT 1080
CCTTTGTGAT GAGCGGTGTG TCACCACATA CAACGATTGT CGTACCTTCC TTGTCTTCTA 1140
AGTGTGATTT CGCCATTTGC ACTGCATGCG CAGTACCGAG TTGTTCCTCT TGAAAACTGT 1200
ATAAAGAACG CTCGCCTAAA TGTCCTTTTA CACTTTCAGC ACCATGTCCT ACGATGGTTA 1260
CAACTTGATC GACACCAGAG CCTTTCACAC TTTCCAATAC ATGTTCGACC ATAGGTTTCC 1320
CAGCAACCTC GTGTAGCACT TTATACTTTT TAGATTTCAT TCTTGTGCCT TTACCTGCTG 1380
CCAAAATTAT CGCGTGTCTT CGCATGAACA TTAACCCCCA TTAAATTTTA CACTTCGTCA 1440
TTATTATAA 1449

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 450 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE: internal
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Arg Arg His Ala Ile Ile Leu Ala Ala Gly Lys Gly Thr Arg Met
1 5 10 15
Lys Ser Lys Lys Tyr Lys Val Leu His Glu Val Ala Gly Lys Pro Met
20 25 30
Val Glu His Val Leu Glu Ser Val Lys Gly Ser Gly Val Asp Gln Val
35 40 45
Val Thr Ile Val Gly His Gly Ala Glu Ser Val Lys Gly His Leu Gly
50 55 60
Glu Arg Ser Leu Tyr Ser Phe Gln Glu Glu Gln Leu Gly Thr Ala His
65 70 75 80
Ala Val Gln Met Ala Lys Ser His Leu Glu Asp Lys Glu Gly Thr Thr
85 90 95
Ile Val Val Cys Gly Asp Thr Pro Leu Ile Thr Lys Glu Thr Leu Glu
100 105 110
Thr Leu Ile Ala His His Glu Asp Ala Asn Ala Gln Ala Thr Val Leu
115 120 125
Ser Ala Ser Ile Gln Gln Pro Tyr Gly Tyr Gly Arg Ile Val Arg Asn
130 135 140

```
Ala Ser Gly Arg Leu Glu Arg Ile Val Glu Glu Lys Asp Ala Thr Gln
145                 150                 155                 160
Ala Glu Lys Asp Ile Asn Glu Ile Ser Ser Gly Ile Phe Ala Phe Asn
                165                 170                 175
Asn Lys Thr Leu Phe Glu Lys Leu Thr Gln Val Lys Asn Asp Asn Ala
            180                 185                 190
Gln Gly Glu Tyr Tyr Leu Pro Asp Val Leu Ser Leu Ile Leu Asn Asp
        195                 200                 205
Gly Gly Ile Val Glu Val Tyr Arg Thr Asn Asp Val Glu Glu Ile Met
    210                 215                 220
Gly Val Asn Asp Arg Val Met Leu Ser Gln Ala Glu Lys Ala Met Gln
225                 230                 235                 240
Arg Arg Thr Asn His Tyr His Met Leu Asn Gly Val Thr Ile Ile Asp
                245                 250                 255
Pro Asp Ser Thr Phe Ile Gly Pro Asp Val Thr Ile Gly Ser Asp Thr
            260                 265                 270
Val Ile Glu Pro Gly Val Arg Ile Asn Gly Arg Thr Glu Ile Gly Glu
        275                 280                 285
Asp Val Val Ile Gly Gln Tyr Ser Glu Ile Asn Asn Ser Thr Ile Glu
    290                 295                 300
Asn Gly Ala Cys Ile Gln Gln Ser Val Val Asn Asp Ala Ser Val Gly
305                 310                 315                 320
Ala Asn Thr Lys Val Gly Pro Phe Ala Gln Leu Arg Pro Gly Ala Gln
                325                 330                 335
Leu Gly Ala Asp Val Lys Val Gly Asn Phe Val Glu Ile Lys Lys Ala
            340                 345                 350
Asp Leu Lys Asp Gly Ala Lys Val Ser His Leu Ser Tyr Ile Gly Asp
        355                 360                 365
Ala Val Ile Gly Glu Arg Thr Asn Ile Gly Cys Gly Thr Ile Thr Val
    370                 375                 380
Asn Tyr Asp Gly Glu Asn Lys Phe Lys Thr Ile Val Gly Lys Asp Ser
385                 390                 395                 400
Phe Val Gly Cys Asn Val Asn Leu Val Ala Pro Val Thr Ile Gly Asp
                405                 410                 415
Asp Val Leu Val Ala Ala Gly Ser Thr Ile Thr Asp Asp Val Pro Asn
            420                 425                 430
Asp Ser Leu Ala Val Ala Arg Ala Arg Gln Thr Thr Lys Glu Gly Tyr
        435                 440                 445
Arg Lys
    450
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

TATATTCATT ATTTAACA 18

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

TGGTGTGCGG AGCGATTA 18

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 452 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO

(v) FRAGMENT TYPE: N-terminal

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
Met Phe Met Arg Arg His Ala Ile Ile Leu Ala Ala Gly Lys Gly Thr
 1               5                  10                  15
Arg Met Lys Ser Lys Lys Tyr Lys Val Leu His Glu Val Ala Gly Lys
            20                  25                  30
Pro Met Val Glu His Val Leu Glu Ser Val Lys Gly Ser Gly Val Asp
        35                  40                  45
Gln Val Val Thr Ile Val Gly His Gly Ala Glu Ser Val Lys Gly His
    50                  55                  60
Leu Gly Glu Arg Ser Leu Tyr Ser Phe Gln Glu Glu Gln Leu Gly Thr
65                  70                  75                  80
Ala His Ala Val Gln Met Ala Lys Ser His Leu Glu Asp Lys Glu Gly
                85                  90                  95
Thr Thr Ile Val Val Cys Gly Asp Thr Pro Leu Ile Thr Lys Glu Thr
            100                 105                 110
Leu Glu Thr Leu Ile Ala His His Glu Asp Ala Asn Ala Gln Ala Thr
        115                 120                 125
Val Leu Ser Ala Ser Ile Gln Gln Pro Tyr Gly Tyr Gly Arg Ile Val
    130                 135                 140
Arg Asn Ala Ser Gly Arg Leu Glu Arg Ile Val Glu Glu Lys Asp Ala
145                 150                 155                 160
Thr Gln Ala Glu Lys Asp Ile Asn Glu Ile Ser Ser Gly Ile Phe Ala
                165                 170                 175
Phe Asn Asn Lys Thr Leu Phe Glu Lys Leu Thr Gln Val Lys Asn Asp
            180                 185                 190
Asn Ala Gln Gly Glu Tyr Tyr Leu Pro Asp Val Leu Ser Leu Ile Leu
        195                 200                 205
Asn Asp Gly Gly Ile Val Glu Val Tyr Arg Thr Asn Asp Val Glu Glu
    210                 215                 220
Ile Met Gly Val Asn Asp Arg Val Met Leu Ser Gln Ala Glu Lys Ala
225                 230                 235                 240
Met Gln Arg Arg Thr Asn His Tyr His Met Leu Asn Gly Val Thr Ile
                245                 250                 255
Ile Asp Pro Asp Ser Thr Phe Ile Gly Pro Asp Val Thr Ile Gly Ser
            260                 265                 270
Asp Thr Val Ile Glu Pro Gly Val Arg Ile Asn Gly Arg Thr Glu Ile
        275                 280                 285
Gly Glu Asp Val Val Ile Gly Gln Tyr Ser Glu Ile Asn Asn Ser Thr
    290                 295                 300
```

```
Ile Glu Asn Gly Ala Cys Ile Gln Gln Ser Val Val Asn Asp Ala Ser
305                 310                 315                 320
Val Gly Ala Asn Thr Lys Val Gly Pro Phe Ala Gln Leu Arg Pro Gly
                325                 330                 335
Ala Gln Leu Gly Ala Asp Val Lys Val Gly Asn Phe Val Glu Ile Lys
            340                 345                 350
Lys Ala Asp Leu Lys Asp Gly Ala Lys Val Ser His Leu Ser Tyr Ile
        355                 360                 365
Gly Asp Ala Val Ile Gly Glu Arg Thr Asn Ile Gly Cys Gly Thr Ile
    370                 375                 380
Thr Val Asn Tyr Asp Gly Glu Asn Lys Phe Lys Thr Ile Val Gly Lys
385                 390                 395                 400
Asp Ser Phe Val Gly Cys Asn Val Asn Leu Val Ala Pro Val Thr Ile
                405                 410                 415
Gly Asp Asp Val Leu Val Ala Ala Gly Ser Thr Ile Thr Asp Asp Val
            420                 425                 430
Pro Asn Asp Ser Leu Ala Val Ala Arg Ala Arg Gln Thr Thr Lys Glu
        435                 440                 445
Gly Tyr Arg Lys
    450
```

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

GCTCCTAAAA GGTTACTCCA CCGGC 25

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description on page 6, line 1.

B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet ☐

Name of depositary institution
National Collection of Industrial and Marine Bacteria Ltd. (NCIMB)

Address of depositary institution *(including postal code and country)*
Aberdeen
Scotland

Date of deposit	Accession Number
11 September 1995	NCIMB 40771

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)* This information is continued on an additional sheet ☐

In respect of those designations in which a European Patent is sought, a sample of the deposited micro-organism will be made available until the publication of the mention of grant of the European Patent or until the date on which the application has been refused or withdrawn, only by the issue of such a sample to an expert nominated by the person requesting the sample.

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later *(specify the general nature of the indications e.g., "Accession Number of Deposit")*

For receiving Office use only	For International Bureau use only
☐ This sheet was received with the international application	☐ This sheet was received by the International Bureau on:
Authorized officer	Authorized officer

Form PCT/RO/134 (July 1992)

What is claimed is:

1. An isolated polynucleotide comprising a member selected from the group consisting of:

(a) a polynucleotide having at least a 70% identity to a polynucleotide encoding a polypeptide comprising amino acids 1 to 450 of SEQ ID NO:2;

(b) a polynucleotide having at least a 70% identity to a polynucleotide encoding a polypeptide comprising amino acids 1 to 452 of SEQ ID NO:5;

(c) a polynucleotide which is complementary to the polynucleotide of (a);

(d) a polynucleotide which is complementary to the polynucleotide of (b); and

(e) a polynucleotide comprising at least 15 sequential bases of the polynucleotide of (a), (b), (c) or (d).

2. The polynucleotide of Claim 1 wherein the polynucleotide is DNA.

3. The polynucleotide of Claim 1 wherein the polynucleotide is RNA.

4. The polynucleotide of Claim 2 comprising a polynucleotide sequence selected from the group consisting of the complementary sequence to nucleotide 54 to 1406 set forth in SEQ ID NO:1, and a complementary sequence to nucleotide 54 to 1412 set forth in SEQ ID NO:1.

5. The polynucleotide of Claim 2 comprising the polynucleotide set forth in SEQ ID NO:1 that encodes novel NAGPU polypeptide.

6. The polynucleotide of Claim 2 which encodes a polypeptide comprising amino acid selected from the group consisting of amino acid 1 to 450 of SEQ ID NO:2 and amino acid 1 to 452 of SEQ ID NO:5.

7. An isolated polynucleotide comprising a member selected from the group consisting of:

(a) a polynucleotide having at least a 70% identity to a polynucleotide encoding the same mature polypeptide expressed by the DNA contained in NCIMB Deposit No. 40794 and having the polynucleotide sequence of SEQ ID NO:1;

(b) a polynucleotide complementary to the polynucleotide of (a); and

(c) a polynucleotide comprising at least 15 bases of the polynucleotide of (a) or (b).

8. A vector comprising the DNA of Claim 2.

9. A host cell comprising the vector of Claim 8.

10. A process for producing a polypeptide comprising: expressing from the host cell of Claim 9 a polypeptide encoded by said DNA.

11. A process for producing a cell which expresses a polypeptide comprising transforming or transfecting the cell with the vector of Claim 8 such that the cell expresses the polypeptide encoded by the cDNA contained in the vector.

12. A polypeptide comprising an amino acid sequence which is at least 70% identical to amino acid selected from the group consisting of 1 to 450 of SEQ ID NO:2 and amino acid 1 to 452 of SEQ ID NO:5.

13. A polypeptide comprising an amino acid sequence selected from the group consisting of the amino acids as set forth in SEQ ID NO:2 and the amino acids as set forth in SEQ ID NO:5.

14. An antibody against the polypeptide of claim 12.

15. An antagonist which inhibits the activity of the polypeptide of claim 12.

16. A method for the treatment of an individual having need of novel NAGPU protein comprising: administering to the individual a therapeutically effective amount of the polypeptide of claim 12.

17. The method of Claim 16 wherein said therapeutically effective amount of the polypeptide is administered by providing to the individual DNA encoding said polypeptide and expressing said polypeptide *in vivo*.

18. A method for the treatment of an individual having need to inhibit novel NAGPU polypeptide comprising: administering to the individual a therapeutically effective amount of the antagonist of Claim 15.

19. A process for diagnosing a disease related to expression of the polypeptide of claim 12 comprising:

determining a nucleic acid sequence encoding said polypeptide.

20. A diagnostic process comprising:

analyzing for the presence of the polypeptide of claim 12 in a sample derived from a host.

21. A method for identifying compounds which bind to and inhibit an activity of the polypeptide of claim 12 comprising:

contacting a cell expressing on the surface thereof a binding means for the polypeptide, said binding means being associated with a second component capable of

providing a detectable signal in response to the binding of a compound to said binding means, with a compound to be screened under conditions to permit binding to the binding; and

determining whether the compound binds to and activates or inhibits the binding by detecting the presence or absence of a signal generated from the interaction of the compound with the binding means.

22. A method for inducing an immunological response in a mammal which comprises inoculating the mammal with novel NAGPU protein, or a fragment or variant thereof, adequate to produce antibody to protect said animal from disease.

23. A method of inducing immunological response in a mammal which comprises, through gene therapy, delivering a gene encoding a novel NAGPU protein fragment or a variant thereof, for expressing novel NAGPU protein, or a fragment or a variant thereof *in vivo* in order to induce an immunological response to produce antibody to protect said animal from disease.

24. An immunological composition comprising a DNA which codes for and expresses a novel NAGPU polynucleotide or protein coded therefrom which, when introduced into a mammal, induces an immunological response in the mammal to a given novel NAGPU polynucleotide or protein coded therefrom.

FIGURE 1A [SEQ ID NO:1]

```
  1  TAATATTACA TAGGATTATC CTTTTATCCT AGCCATTTTA AATACGTAAA

 51  TGATTATTTC CTATATCCTT CTTTTGTTGT TTGTCTTGCT CTTGCCACAG

101  CTAAACTGTC ATTTGGTACG TCATCTGTGA TTGTGGAACC AGCTGCCACC

151  AATACATCAT CACCAATTGT TACAGGTGCT ACTAAATTAA CATTGCAACC

201  TACAAATGAA TCTTTGCCGA CGATAGTTTT AAATTTATTT TCACCATCAT

251  AGTTAACTGT AATCGTTCCG CAACCAATAT TAGTACGTTC GCCAATTACA

301  GCATCGCCAA TATAACTTAA ATGTGAAACC TTGGCACCAT CTTTAAGATC

351  TGCTTTTTTA ATTTCTACAA AATTTCCAAC CTTAACATCT GCACCTAATT

401  GCGCGCCTGG TCTCAATTGC GCAAACGGTC CGACCTTAGT ATTAGCTCCT

451  ACGCTAGCAT CATTAACAAC AGACTGTTGA ATACATGCAC CATTTTCAAT

501  CGTACTATTG TTAATTTCAG AGTACTGACC AATAACAACA TCTTCGCCAA

551  TTTCTGTACG ACCATTAATT CGTACGCCTG GTTCAATGAC TGTATCACTA

601  CCAATTGTAA CGTCTGGACC AATAAAAGTG CTGTCAGGAT CGATGATTGT
```

FIGURE 1B [SEQ ID NO:1]

```
 651  CACACCATTT AGCATGTGAT AATGATTCGT ACGACGTTGC ATCGCCTTCT
 701  CAGCCTGACT AAGCATTACA CGATCATTTA CACCCATGAT TTCTTCAACA
 751  TCATTGGTAC GATACACTTC TACGATGCCG CCATCATTTA AAATTAACGA
 801  CAATACATCA GGGAGGTAAT ATTCACCTTG CGCATTATCA TTTTTCACTT
 851  GTGTTAATTT TTCAAACAAC GTTTTATTAT TAAACGCAAA AATACCTGAA
 901  CTAATTTCAT TAATATCCTT TTCAGCTTGC GTTGCATCTT TCTCTTCAAC
 951  TATGCGTTCT AAACGACCTG ACGCATTTCG AACGATTCTT CCGTATCCAT
1001  ATGGTTGTTG AATCGATGCA GATAATACAG TTGCTTGAGC ATTAGCATCC
1051  TCATGATGCG CAATCAATGT TTCTAATGTT TCCTTTGTGA TGAGCGGTGT
1101  GTCACCACAT ACAACGATTG TCGTACCTTC CTTGTCTTCT AAGTGTGATT
1151  TCGCCATTTG CACTGCATGC GCAGTACCGA GTTGTTCCTC TTGAAAACTG
1201  TATAAAGAAC GCTCGCCTAA ATGTCCTTTT ACACTTTCAG CACCATGTCC
1251  TACGATGGTT ACAACTTGAT CGACACCAGA GCCTTTCACA CTTTCCAATA
```

FIGURE 1C [SEQ ID NO:1]

1301 CATGTTCGAC CATAGGTTTC CCAGCAACCT CGTGTAGCAC TTTATACTTT

1351 TTAGATTTCA TTCTTGTGCC TTTACCTGCT GCCAAAATTA TCGCGTGTCT

1401 TCG**CAT**GAA**C AT**TAACCCCCC ATTAAATTTT ACACTTCGTC ATTATTATAA

FIGURE 2 [SEQ ID NO:2]

```
  1  MRRHAIILAA GKGTRMKSKK YKVLHEVAGK PMVEHVLESV KGSGVDQVVT

 51  IVGHGAESVK GHLGERSLYS FQEEQLGTAH AVQMAKSHLE DKEGTTIVVC

101  GDTPLITKET LETLIAHHED ANAQATVLSA SIQQPYGYGR IVRNASGRLE

151  RIVEEKDATQ AEKDINEISS GIFAFNNKTL FEKLTQVKND NAQGEYYLPD

201  VLSLILNDGG IVEVYRTNDV EEIMGVNDRV MLSQAEKAMQ RRTNHYHMLN

251  GVTIIDPDST FIGPDVTIGS DTVIEPGVRI NGRTEIGEDV VIGQYSEINN

301  STIENGACIQ QSVVNDASVG ANTKVGPFAQ LRPGAQLGAD VKVGNFVEIK

351  KADLKDGAKV SHLSYIGDAV IGERTNIGCG TITVNYDGEN KFKTIVGKDS

401  FVGCNVNLVA PVTIGDDVLV AAGSTITDDV PNDSLAVARA RQTTKEGYRK
```

FIGURE 3 [SEQ ID NO:3]

tatattcatt atttaaca

[SEQ ID NO:4]

tggtgtgcgg agcgatta

FIGURE 4 [SEQ ID NO:5]

```
  1  MFMRRHAIIL AAGKGTRMKS KKYKVLHEVA GKPMVEHVLE SVKGSGVDQV
 51  VTIVGHGAES VKGHLGERSL YSFQEEQLGT AHAVQMAKSH LEDKEGTTIV
101  VCGDTPLITK ETLETLIAHH EDANAQATVL SASIQQPYGY GRIVRNASGR
151  LERIVEEKDA TQAEKDINEI SSGIFAFNNK TLFEKLTQVK NDNAQGEYYL
201  PDVLSLILND GGIVEVYRTN DVEEIMGVND RVMLSQAEKA MQRRTNHYHM
251  LNGVTIIDPD STFIGPDVTI GSDTVIEPGV RINGRTEIGE DVVIGQYSEI
301  NNSTIENGAC IQQSVVNDAS VGANTKVGPF AQLRPGAQLG ADVKVGNFVE
351  IKKADLKDGA KVSHLSYIGD AVIGERTNIG CGTITVNYDG ENKFKTIVGK
401  DSFVGCNVNL VAPVTIGDDV LVAAGSTITD DVPNDSLAVA RARQTTKEGY
451  RK
```

INTERNATIONAL SEARCH REPORT

Inte nal Application No
PCT/GB 96/03047

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/54 C12N1/21 C12N9/12 C12Q1/68 C07K16/40
A61K48/00 A61K38/43 G01N33/53

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N C07K C12Q A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	JOURNAL OF BACTERIOLOGY, vol. 177, no. 23, December 1993, pages 6902-6909, XP000647399 ULLRICH, J. AND VAN PUTTEN, J.P.M.: "IDENTIFICATION OF THE GONOCOCCAL glmU GENE ENCODING THE ENZYME N-ACETYLGLUCOSAMINE 1-PHOSPHATE URIDYLTRANSFERASE INVOLVED IN THE SYNTHESIS OF UDP-GlcNAc" see the whole document --- -/--	1,2,7-11

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

2

Date of the actual completion of the international search	Date of mailing of the international search report
14 March 1997	25. 03. 97

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+ 31-70) 340-3016	Holtorf, S

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Inte nal Application No
PCT/GB 96/03047

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	JOURNAL OF BACTERIOLOGY, vol. 175, no. 19, October 1993, pages 6150-6157, XP000647398 MENGIN-LECREULX, D. AND VAN HEIJENOORT, J.: "IDENTIFICATION OF THE glmU GNE ENCODING N-ACETYLGLUCOSAMINE-1-PHOSPHATE URIDYLTRANSFERASE IN ESCHERICHIA COLI" cited in the application see the whole document	1,2,7-11
X	MOLECULAR AND GENERAL GENETICS, vol. 218, 1989, pages 565-571, XP002027590 NILSSON, D., ET AL . : "PRIMARY STRUCTURE OF THE tms AND prs GENES OF BACILLUS SUBTILIS" see the whole document	1,2,7,8
X	EP 0 096 547 A (SEITETSU KAGAKU CO LTD) 21 December 1983	12,13
Y	see page 4	1,2,4-9
Y	EMBL SEQUENCE DATA LIBRARY, 17 January 1995, HEIDELBERG, GERMANY, XP002027591 HOVE-JENSEN,B.: "BACILLUS CALDOLYTICUS PRS GENE ENCODING PHOPHORIBOSYL-DIPHOSPHATE SYNTHASE" ACCESSION NO. X83708	1,2,4-9
A	SCIENCE, vol. 259, January 1993, pages 686-688, XP002027592 MAHAN, J.M., ET AL . : "SELECTION OF BACTERIAL VIRULENCE GENES THAT ARE SPECIFICALLY INDUCED IN HOST TISSUES" cited in the application see the whole document	1-24
A	WO 93 24136 A (TERMAN DAVID S ;STONE JAY L (US)) 9 December 1993 see the whole document	23
A	WO 94 27435 A (LIFE TECHNOLOGIES INC) 8 December 1994 see the whole document	24

2

INTERNATIONAL SEARCH REPORT

ernational application No.
PCT/GB 96/03047

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.: 16-19,22,23
because they relate to subject matter not required to be searched by this Authority, namely:

Remark: Although these claims are directed to a method of treatment of (diagnostic method practised on) the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

2. [ ] Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searches without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [ ] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

Information on patent family members

Inv onal Application No
PCT/GB 96/03047

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0096547 A	21-12-83	JP 1346222 C	13-11-86
		JP 58212795 A	10-12-83
		JP 61011597 B	03-04-86
		JP 1374014 C	07-04-87
		JP 58212799 A	10-12-83
		JP 61039039 B	02-09-86
		JP 58212796 A	10-12-83
		JP 1050397 B	30-10-89
		JP 1565405 C	25-06-90
		JP 58212797 A	10-12-83
		DE 3376020 A	21-04-88
		EP 0223263 A	27-05-87
		EP 0223264 A	27-05-87
		EP 0220750 A	06-05-87
		US 4569909 A	11-02-86
		US 4604349 A	05-08-86
WO 9324136 A	09-12-93	AU 4526593 A	30-12-93
		WO 9420124 A	15-09-94
WO 9427435 A	08-12-94	EP 0702516 A	27-03-96
		JP 9500013 T	07-01-97